



# Study on Vegetation Classification Based on Spectral Knowledge Base

Peng Liu, Jingcheng Zhang<sup>(✉)</sup>, Bin Wang, Xuexue Zhang,  
and Kaihua Wu

College of Life Information Science and Instrument Engineering,  
Hangzhou Dianzi University, Hangzhou 310018, Zhejiang, China  
zzwliupeng@163.com, zhangjc\_rs@163.com,  
wangb\_rs@163.com, zhangxx\_rs@163.com,  
wukaihua@hdu.edu.cn

**Abstract.** A framework about spectral based vegetation classification was proposed, which serves as a core methodology of the vegetation spectral knowledge base. The hyperspectral reflectances of 13 types of plants were measured by an ASD FieldSpec 4 spectroradiometer. Two forms of spectral features were used for representing the key spectral characteristics of plants, including Vegetation index (VI) and spectral shape features. Based on these spectral features, a sensitivity analysis was performed to identify the most important features for establishing the classifier. The analysis of variance (ANOVA) and the cross-correlation analysis were applied to derive the sensitivity of features and remove features that have high correlations. Then, a classification method for differentiating plants was established by coupling some spectral similarity measures (e.g., ED) with some classification methods (e.g., BPANN and SVM). The results of discrimination analysis showed that a highest accuracy was produced by SVM with the OAA over 99% when using 7 sensitive VIs. The results suggested the framework about spectral based vegetation classification can form a basis for spectral knowledge base and application technology and further achieve a wide range of plant classification based on remote sensing.

**Keywords:** Vegetation classification · Hyperspectral · Feature extraction  
Classification algorithm

## 1 Introduction

Recent advances in Hyperspectral provide opportunities to map plant species and vegetation at various scales and resolutions. Establish a spectral knowledge base is an effective management tool. Among them, the study of vegetation classification methods based on hyperspectral data is an important part of plant spectrum library. Over the last decade, vegetation canopy spectral reflectance has been successfully used in discriminating plant species (Schmidt and Skidmore 2003; Pu 2009; Allard et al. 2011; Peñuelas 1995a). Some detailed changes in spectral curves of hyperspectral data can be detected by spectral feature selection and extraction methods such as continuum

removal or derivative analysis. (Schmidt and Skidmore 2003; Abdel-Rahman et al. 2010). Gong et al. (1997) and Pu (2009) found that the first derivative of the tree spectrum can significantly improve the accuracy of the identification of the six species of common conifers in northern California and the 11 species of urban trees in Tampa, Florida. Prospero et al. (2014) used 47 spectral variables to classify 46 plant species in tropical wetland. A series of methods were used to select and extract features. Then a set of algorithms were used to build the classification model. Pu (2011) utilized a stepwise masking system to process the high-resolution IKONOS images and to identify and map urban forest tree species/groups. Bue et al. (2015) proposed a method of combining Multiple Endmember Spectral Mixture Analysis (MESMA) and Multi-class Discriminant Analysis (MDA) for classifying spectrally-similar species. In a case study for mapping urban trees, Pu and Landry (2012) found that the WV2 imagery produced a higher accuracy than the IKONOS imagery according to an independent validation. Zeng et al. (2017) chooses the wetland of Poyang Lake as the study area, and with the help of ASD FieldSpec4, it gets the reflecting curve of six advantage kinds of vegetation. Based on the pretreatment of original spectral data, this research uses a series of methods to analyze and compare the spectral curve, including derivative, Log (1/R), continuum removing and so on. Then, it uses the characteristic parameters of spectral to classify and extract the plant. Yu et al. (2017) used spectral derivative method and vegetation index method to construct spectral features. The artificial neural network method and factor analysis method were used to classify and extract typical vegetation.

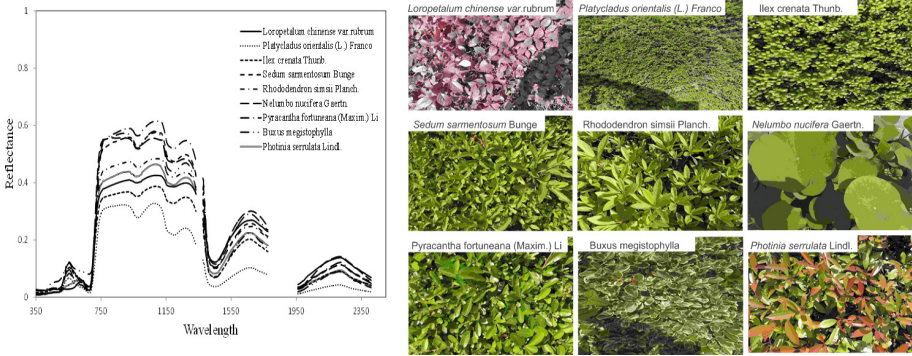
The spectral based vegetation classification is one of the core technologies for establishing plant spectrum library. Based on hyperspectral measurements of a number of plant species, this paper focuses on the extraction and selection of spectral features for vegetation classification. Besides, combining with some pattern recognition algorithms, a framework for spectral based vegetation classification was established and evaluated.

## 2 Materials and Methods

### 2.1 Experimental Design and Data Acquisition

In this study, plant hyperspectral measurements were made in May 2017 in Hangzhou (Lon 120.34°, Lat 30.31°), where consisting a variety of plants. The plant canopy spectrometer was measured using an ASD Fieldspec FR2500 spectrometer with a sampling interval of 1.4 nm (350 to 1000 nm) and 2 nm (in the range of 1000 to 2500 nm). All spectral measurements are made in windless weather within 10:00 ~ 14:00. The sensor was placed at 60 cm above the plants' canopy. The spectral data is calibrated against a white reference panel to convert from the radiance signal to spectral reflectance. A total of 20 recordings were made and were then averaged to get one canopy spectrum. Per plant, 10-25 measurements were made in the field of view, and 13 species of plant hyperspectral data (n = 324) was Measured, including *Loropetalum chinense var. rubrum*, *Platyclusus orientalis (L.) Franco*, *Ilex crenata* Thunb., *Sedum sarmentosum* Bunge, *Rhododendron simsii* Planch., *Nelumbo nucifera* Gaertn.,

*Pyracantha fortuneana* (Maxim.) Li, *Buxus megistophylla*, *Ligustrum vicaryi*, *Paspalum thunbergii* Kunth ex Steud., *Mahonia fortune* (Lindl.) Fedde, *Photinia serrulata* Lindl. and *Buxus sinica* (Rehd. et Wils.) Cheng (Fig. 1).



**Fig. 1.** Example of hyperspectral curves of different plants and their canopy pictures

### 2.2 Features Extraction

**Vegetation Indices.** In this paper, 13 classic vegetation indices (VIs) are purposely selected for plants classification (Table 1). Some of the VIs are sensitive to pigments variation, some of the VIs are sensitive to variation of leaf area or biomass, while some others are sensitive to the plants’ water contents. The different mechanisms of these VIs are important for indicating the spectral difference among plants. Besides, a sensitivity analysis by the analysis of variance (ANOVA) was performed to identify the most important VIs for establishing the classification model. In addition, a cross-correlation analysis were applied to eliminate features with relatively high information redundancy.

**Table 1.** The 13 vegetation indices extracted from hyperspectral data

Spectral index	Definition	Reference
<b>EVI</b> , Enhanced VI	$2.5(R_{NIR} - R_{red}) / (R_{NIR} + 6R_{red} - 7.5R_{blue} + 1)$	Huete et al. 2002
<b>LAIDI</b> , LAI Determining Index	$R_{1250} / R_{1050}$	Delalieux et al. 2008
<b>NDVI</b> , Normalized Difference Vegetation Index	$(R_{NIR} - R_{red}) / (R_{NIR} + R_{red})$	Rouse et al. 1973
<b>NRI</b> , Normalized Ratio Index	$(R_{874} - R_{1225}) / (R_{874} + R_{1225})$	Koppe et al. 2010
<b>sLAIDI</b> , Normalization or Standard of the LAIDI	$S(R_{1050} - R_{1250}) / (R_{1050} + R_{1250})$ , where $S = 5$	Delalieux et al. 2008

(continued)

**Table 1.** (continued)

Spectral index	Definition	Reference
<b>ACI</b> , Anthocyanin Content Index	$R_{\text{green}}/R_{\text{NIR}}$	Van den Berg and Perkins2005
<b>NPQI</b> , Normalized Phaeophytinization Index	$(R_{415} - R_{435})/(R_{415} + R_{435})$	Barnes et al. 1992; Peñuelas et al. 1995b
<b>PRI</b> , Photochemical/Physiological Reflectance Index	$(R_{531} - R_{570})/(R_{531} + R_{570})$	Gamon et al. 1992
<b>DSWI</b> , Disease Water Stress Index	$(R_{802} + R_{547})/(R_{1657} + R_{682})$	Galvão et al. 2005
<b>RVI<sub>hyp</sub></b> , Hyperspectral Ratio VI	$R_{1088}/R_{1148}$	Schlerf et al. 2005
<b>SIWSI</b> , Shortwave Infrared Water Stress Index	$(R_{860} - R_{1640})/(R_{860} + R_{1640})$	Fensholt and Sandholt 2003
<b>WI</b> , Water Index	$R_{900}/R_{970}$	Peñuelas et al. 1997
<b>RVSI</b> , Red-Edge Vegetation Stress Index	$[(R_{712} + R_{752})/2] - R_{732}$	Merton and Huntington 1999

**Spectral Shape Feature.** Besides the VIs, some spectral shape feature including the first-order derivative features and continuum removal (CRM) features are also included. The derivative spectrum is the normalized spectral difference of two adjacent narrow-bands with their wavelength interval. Spectral derivative analysis can partially eliminate the effects of atmospheric effects, vegetation and environmental background (shadow, soil, etc.) which thus emphasize some essential characteristics of the plant (Demetriades-Shah et al. 1990; Tsai and Philpot 1998). The first derivative spectra can be calculated by:

$$\rho'(\lambda_i) \approx [\rho(\lambda_{i+1}) - \rho(\lambda_{i-1})]/\Delta\lambda \quad (1)$$

where,  $\rho'(\lambda_i)$  is the first derivative.;  $\rho(\lambda_i)$  is the reflectance at a wavelength (band)  $i$ ;  $\Delta\lambda$  is the wavelength interval between  $\lambda_{i+1}$  and  $\lambda_{i-1}$  and equals twice bandwidth for this case. The continuum removal technique provided a quantitative measure of absorption features in plants' spectra (Fig. 2). Thereby, a total of 12 spectral shape features, including 9 first order differential features of red edge, blue edge and yellow edge, and the depth, width and area of CRM in bands range from 530 nm to 770 nm were extracted from the original spectra (Table 2).

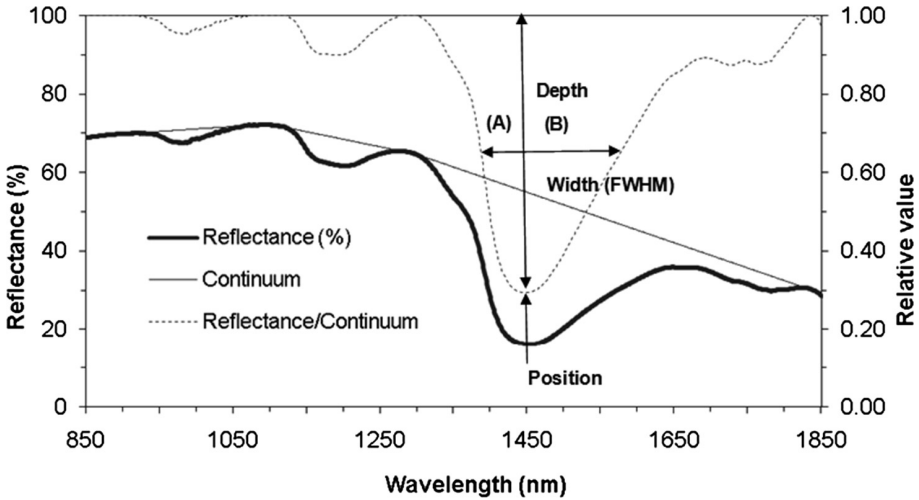


Fig. 2. Illustrated of CRM

Table 2. 12 spectral shape features

	Position	Band range (nm)	Features
First order differential features	Blue edge	490–540	Maximum differential value (BMV)
			Position of the maximum differential value (BPMV)
			Sum of differential values (BSV)
	Yellow edge	540–620	Maximum differential value (YMV)
			Position of the maximum differential value (YPMV)
			Sum of differential values (YSV)
	Red edge	660–780	Maximum differential value (RMV)
			Position of the maximum differential value (RPMV)
			Sum of differential values (RSV)
CRM features	NIR	530–770	Depth
			Width
			Area

### 2.3 Classification Algorithm

To achieve the spectral classification among different plants, three classification algorithms are tested, including Euclidian distance (ED), Support vector machines (SVM) and Artificial neural network (ANN). The Euclidian distance (ED) provides a measure of distance between two pixels or between a reference spectrum and a test spectrum in the  $L$ -dimensional feature space:

$$ED(\rho_t, \rho_r) = \left[ \sum_{i=1}^L (\rho_{ti} - \rho_{ri})^2 \right]^{\frac{1}{2}} \quad (2)$$

Where,  $\rho_r = [\rho_{r1}, \rho_{r2}, \dots, \rho_{rL}]^T$  and  $\rho_t = [\rho_{t1}, \rho_{t2}, \dots, \rho_{tL}]^T$  are assumed as the reference spectrum (a laboratory or pixel spectrum known to characterize a target of interest) and the test spectrum, respectively.  $L$  represents the spectral dimensionality and equals to the number of bands of hyperspectral data (Kong et al. 2010). The SVM is a new type of classifiers that can efficiently overcome the Hughes phenomenon by directly seeking a separating surface (hyperplane) through an optimization procedure. The ANN has been proven that it works well even with small training sample size. The general principal of the ANN is demonstrated in Fig. 3.

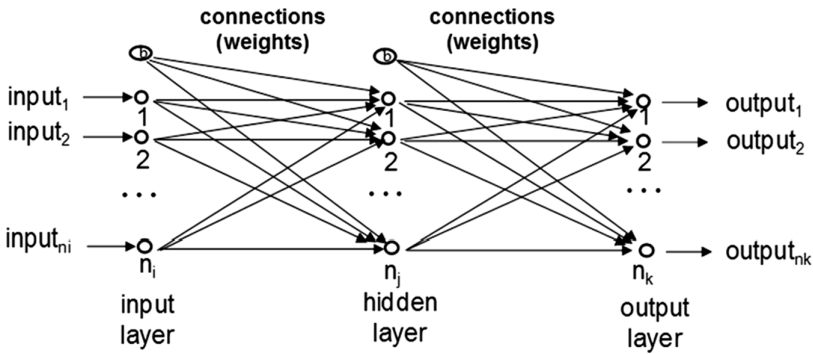
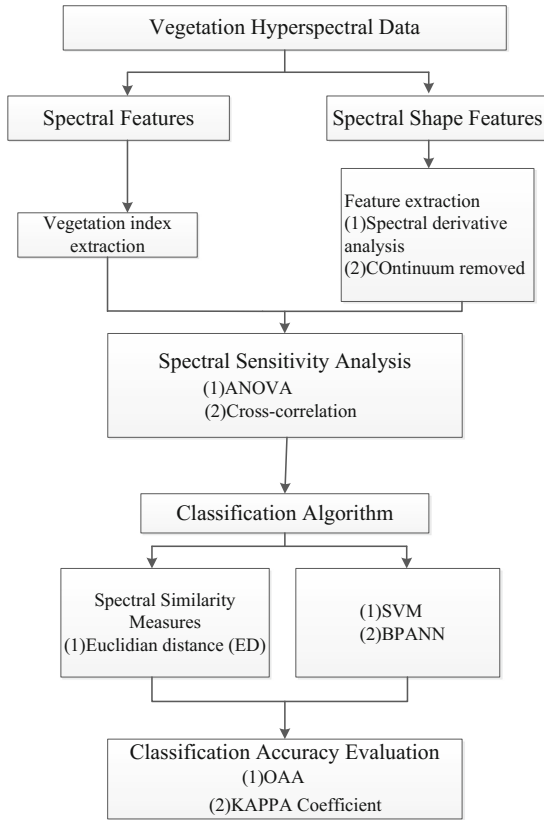


Fig. 3. Neural network structure

Based on the two types of spectral features, the plants classification models were established according to ED, SVM and ANN, respectively. The models were trained with 60% of the data ( $n = 195$ ), whereas were validated against the left 40% of the data ( $n = 129$ ). The accuracies of the classification are evaluated according to the overall accuracy (OAA) and kappa coefficient (Congalton and Mead 1983; Story and Congalton 1986). A flowchart of the entire analysis is demonstrated in Fig. 4.



**Fig. 4.** Experimental flow graph

### 3 Results

#### 3.1 Classification with VIs

According to the sensitivity analysis, a total of 13 VIs passed the ANOVA with p-value <0.01. Then, after the cross correlation check, 7 VIs were retained, including GI, NRI, ACI, RVSI, NDVI, NPQI, and TVI. The means of the 7 VIs were compared among all types of plants as shown in Fig. 5. The classification accuracies of ED, SVM and ANN were shown in Table 3.

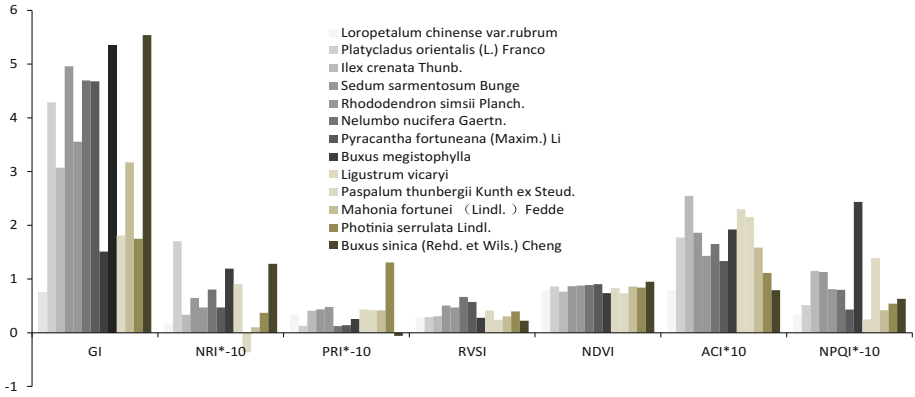


Fig. 5. Illustrates statistics of VIs

Table 3. Confusion matrix created using VIs

	OAA	KAPPA
ED	0.92	0.91
SVM	0.99	0.99
BPANN	0.84	0.82

### 3.2 Classification with Spectral Shape Features

According to the sensitivity analysis, a total of 12 spectral shape features passed the ANOVA with p-value <0.01. Then, after the cross correlation check, 7 spectral shape features were retained, including BMV, BPMV, YPMV, YPMV, RMV, Depth and Width. The means of the spectral shape features were compared among all types of plants as shown in Fig. 6. The classification accuracies of ED, SVM and ANN were shown in Table 4.

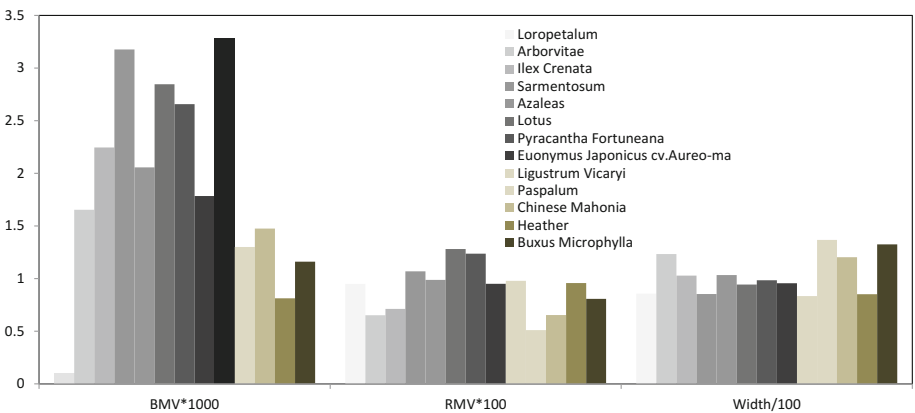


Fig. 6. Illustrates some statistics of spectral shape features



**Table 4.** Confusion matrix created using spectral shape features

	OAA	KAPPA
ED	0.54	0.50
SVM	0.50	0.46
BPANN	0.17	0.09

## 4 Discussions

In this comprehensive study by using hyperspectral data for classifying 13 plant species, it is encouraging that a satisfactory accuracy can be achieved by this spectral classification. Based on the sensitive VIs, the OAA that was produced by all three classification algorithms were over 84%. However, the OAA of the classification models that were based on spectral shape feature were all below 55%, which is not acceptable for application. Comparing with spectral shape features, the significant difference of VIs among plants may account for its higher classification accuracy (Figs. 5 and 6). Among the three classification algorithms, the highest accuracy was produced by SVM with the OAA over 99% whereas the lowest accuracy was produced by ANN. The classification ability based on spectral shape features is weak, and the selection and extraction of more sensitive features are to be further explored.

## 5 Concluding Remarks

In this study, hyperspectral data was used to classify plant species. (1) A total of 7 VIs and 7 spectral shape features were identified as most suitable for plants classification in our case. (2) Based on VIs and SVM algorithm, the highest classification accuracy can be achieved with OAA = 0.99. The results suggested the great potential of hyperspectral data in plants classification. The framework about spectral based vegetation classification can form a basis for spectral knowledge base and application technology and further achieve a wide range of plant classification based on remote sensing.

**Acknowledgements.** This work was supported by Zhejiang public welfare programme of agriculture technology (2016C32087), National Natural Science Foundation of China (41671415; 41601461) and Graduate Scientific Research Foundation of Hangzhou Dianzi University (CXJJ2017068).

## References

- Schmidt, K.S., Skidmore, A.K.: Spectral discrimination of vegetation types in a coastal wetland. *Remote Sens. Environ.* **85**, 92–108 (2003)
- Pu, R.: Broadleaf species recognition with in situ hyperspectral data. *Int. J. Remote. Sensing.* **30** (11), 2759–2779 (2009)
- Allard, D., D'Or, D., Froidevaux, R.: An efficient maximum entropy approach for categorical variable prediction. *Eur. J. Soil Sci.* **62**, 381–393 (2011)

- Peñuelas, J., Baret, F., Filella, I.: Semi-empirical indices to assess carotenoids/chlorophyll, a ratio from leaf spectral reflectance. *Photosynthetica* **31**, 221–230 (1995a)
- Abdel-Rahman, E.M., Ahmed, F.B., Van den Berg, M.: Estimation of sugarcane leaf nitrogen concentration using in situ spectroscopy. *Int. J. Appl. Earth Obs. Geoinf.* **12**, S52–S57 (2010)
- Gong, P., Pu, R., Yu, B.: Conifer species recognition: An exploratory analysis of in situ hyperspectral data. *Remote Sens. Environ.* **62**, 189–200 (1997)
- Prospere, K., McLaren, K., Wilson, B.: Plant species discrimination in a tropical wetland using in situ hyperspectral data. *J. Remote Sens.* **6**(9), 8494–8523 (2014)
- Pu, R.: Mapping urban forest tree species using IKONOS imagery: preliminary results. *J. Environ. Monit. Assess.* **172**(1–4), 199–214 (2011)
- Bue, B.D., Thompson, D.R., Sellar, R.G., Podest, E.V., Eastwood, M.L., Helmlinger, M.C., et al.: Leveraging in-scene spectra for vegetation species discrimination with mesma-mda. *ISPRS J. Photogramm. Remote Sens.* **108**, 33–48 (2015)
- Pu, R., Landry, S.: A comparative analysis of high spatial resolution ikonos and worldview-2 imagery for mapping urban tree species. *Remote Sens. Environ.* **124**(9), 516–533 (2012)
- Zeng, S., Kuang, R., Xiao, Y., Zhao, Z.: Measured hyperspectral data classification of poyang lake wetland vegetation. *Remote Sens. Inf.* **32**(5), 75–81 (2017)
- Yu, J., Li, X., Zhang, Q., Shi, H., Xue, J., Chu, J.: Typical vegetation classification of Taihu lakeside based on measured hyperspectral data. *Jiangsu Agric. Sci.* **45**(5), 240–244 (2017)
- Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X., Ferreira, L.G.: Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sens. Environ.* **83**, 195–213 (2002)
- Delalieux, S., Somers, B., Hereijgers, S., Verstraeten, W.W., Keulemans, W., Coppin, P.: A near-infrared narrow-waveband ratio to determine Leaf Area Index in orchards. *Remote Sens. Environ.* **112**, 3762–3772 (2008)
- Rouse, J.W., Haas, R.H., Schell, J.A., Deering, D.W.: Monitoring vegetation systems in the great plains with ERTS. In: *Proceedings Third ERTS Symposium*, vol. 1, 48–62 (1973)
- Koppe, W., Li, F., Gnyp, M.L., Miao, Y., Jia, L., Chen, X., et al.: Evaluating multispectral and hyperspectral satellite remote sensing data for estimating winter wheat growth parameters at regional scale in the north china plain. *Photogramm.-Fernerkund. – Geoinf.* **3**, 167–178 (2010)
- Berg, A.K.V.D., Perkins, T.D.: Non-destructive estimation of anthocyanin content in autumn sugar maple leaves. *Hortic. Sci.* **40**(3), 685–686 (2005)
- Barnes, J.D., Balaguer, L., Manrique, E., Elvira, S., Davison, A.W.: A reappraisal of the use of DMSO for the extraction and determination of chlorophylls a and b in lichens and higher plants. *Environ. Exp. Bot.* **32**, 85–100 (1992)
- Peñuelas, J., Filella, I., Lloret, P., Muñoz, F., Vilajeliu, M.: Reflectance assessment of mite effects on apple trees. *Int. J. Remote Sens.* **16**, 2727–2733 (1995b)
- Gamon, J.A., Penuelas, J., Field, C.B.: A narrow-waveband spectral index that tracks diurnal changes in photosynthetic efficiency. *Remote Sens. Environ.* **41**, 35–44 (1992)
- Galvão, L.S., Formaggio, A.R., Tisot, D.A.: Discrimination of sugarcane varieties in Southeastern Brazil with EO-1 hyperion data. *Remote Sens. Environ.* **94**, 523–534 (2005)
- Schlerf, M., Atzberger, C., Hill, J.: Remote sensing of forest biophysical variables using HyMap imaging spectrometer data. *Remote Sens. Environ.* **95**, 177–194 (2005)
- Fensholt, R., Sandholt, I.: Derivation of a shortwave infrared water stress index from MODIS near-and shortwave infrared data in a semiarid environment. *Remote Sens. Environ.* **87**(1), 111–121 (2003)
- Peñuelas, J., Piñol, J., Ogaya, R., Filella, I.: Estimation of plant water concentration by the reflectance water index WI (R900/R970). *Int. J. Remote Sens.* **18**, 2869–2875 (1997)

- Merton, R., Huntington, J.: Early simulation of the ARIES-1 satellite sensor for multi-temporal vegetation research derived from AVIRIS. In: *Summaries of the Eight JPL Airborne Earth Science Workshop*, 9–11 February, pp. 299–307. JPL Publication 99-17, Pasadena (1999)
- Demetriades-Shah, T.H., Steven, M.D., Clark, J.A.: High-resolution derivative spectra in remote-sensing. *Remote Sens. Environ.* **33**(1), 55–64 (1990)
- Tsai, F., Philpot, W.: Derivative analysis of hyperspectral data. *Remote Sens. Environ.* **66**(1), 41–51 (1998)
- Kong, X., Shu, N., Huang, W., Fu, J.: The research on effectiveness of spectral similarity measures for hyperspectral image. In: *IEEE 2010 3rd International Congress on Image and Signal Processing (CISP2010)*, pp. 2269–2273 (2010)
- Congalton, R.G., Mead, R.A.: A quantitative method to test for consistency and correctness in photointerpretation. *Photogramm. Eng. Remote. Sens.* **49**(1), 69–74 (1983)
- Story, M., Congalton, R.G.: Accuracy assessment: A user's perspective. *Photogramm. Eng. Remote. Sens.* **52**(3), 397–399 (1986)